

Урок 3. Парсинг данных. HTML, BeautifulSoup

Необходимо собрать информацию о вакансиях на вводимую должность (используем input или через аргументы получаем должность) с сайтов HH(обязательно) и/или Superjob(по желанию). Приложение должно анализировать несколько страниц сайта (также вводим через input или аргументы). Получившийся список должен содержать в себе минимум:

Наименование вакансии. Предлагаемую зарплату (разносим в три поля: минимальная и максимальная и валюта. цифры преобразуем к цифрам). Ссылку на саму вакансию. Сайт, откуда собрана вакансия.

```
In [2]: from bs4 import BeautifulSoup as bs
import requests
import pandas as pd
```

```
In [3]: headers = {'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/88.0.4399.24 Safari/537.36'}
url = 'https://hh.ru/search/vacancy'
```

```
In [32]: def get_html(url, params=''):
response = requests.get(url, headers=headers, params=params)
return response
```

```
In [36]: def hh_content(response):
dom = bs(response, 'lxml')
items = dom.find_all('div', class_='vacancy-serp-item__layout')
hh_vacancy = []

for item in items:
    hh_vacancy.append(
        {
            'site': 'HeadHunter', # Название сайта
            'title': item.find('h3', class_='bloko-header-section-3').get_text(),
            'link': item.find('a', class_='serp-item__title').get('href'), # Ссылка на вакансию
            'salary': item.find('span', class_='bloko-header-section-3').get_text(), # Зарплата
            'organization': item.find('div', class_='vacancy-serp-item-company').get_text()
        }
    )
for i in hh_vacancy:
    try:
        i['salary'] = i['salary'].text
    except:
        i['salary'] = None
    if i['salary']:
        salary_list = i['salary'].split(' ')
        if salary_list[0] == 'от':
            i['salary_min'] = salary_list[1]
            i['salary_max'] = None
        elif salary_list[0] == 'до':
            i['salary_min'] = None
            i['salary_max'] = salary_list[1]
```

```
        else:
            i['salary_min'] = salary_list[0]
            i['salary_max'] = salary_list[2]
            i['salary_currency'] = salary_list[-1]
        else:
            i['salary_min'] = None
            i['salary_max'] = None
            i['salary_currency'] = None
        i.pop('salary')
    return hh_vacancy
```

```
In [37]: def parser_hh():
    post = str(input('Введите название вакансии для парсинга: '))
    pages = int(input('Количество страниц для парсинга: '))
    html = get_html(url)
    if html.status_code == 200:
        vacancy = []
        for page in range(1, pages + 1):
            print(f'Парсим страницу {page}')
            html = get_html(url, params={'text': post, 'page': page})
            vacancy.extend(hh_content(html.text))
        hh_result = pd.DataFrame(vacancy)
    else:
        hh_result = html.status_code
    return hh_result
```

```
In [38]: df = parser_hh()
df
```

Введите название вакансии для парсинга: водитель
Количество страниц для парсинга: 1
Парсим страницу 1

Out[38]:

	site	title	link	organiza
0	HeadHunter	Водитель категории E	https://adsrv.hh.ru/click?b=329606&c=30&meta=D...	МАГНИТ, Розничная сетьС
1	HeadHunter	Водитель на самовал	https://adsrv.hh.ru/click?b=367192&c=30&meta=6...	ООО ТСКТ
2	HeadHunter	Водитель-курьер на автомобиле компании (юг гор...	https://voronezh.hh.ru/vacancy/76640105?from=v...	Орешкофф.рфСа Петерф
3	HeadHunter	Водитель-курьер на личном автомобиле	https://voronezh.hh.ru/vacancy/71302355?from=v...	СберЛогистикаНовосиб
4	HeadHunter	Курьер-водитель в Москве на доставку легких за...	https://voronezh.hh.ru/vacancy/77051322?from=v...	ООО КурьеристМос
5	HeadHunter	Водитель Категории D	https://voronezh.hh.ru/vacancy/76947525?from=v...	ООО Аэросиб СВиД
6	HeadHunter	Водитель на автомобиле компании	https://voronezh.hh.ru/vacancy/72155946?from=v...	LamodaМосква, Дмитров и ел
7	HeadHunter	Водитель специального грузового автомобиля мус...	https://voronezh.hh.ru/vacancy/66792562?from=v...	УП Филиал №1 Унитарн предприятия Вторичнь
8	HeadHunter	Водитель-международник	https://voronezh.hh.ru/vacancy/76084101?from=v...	ООО ИНТЕРТРАНСАВТОМ
9	HeadHunter	Водитель-международник (стажер)	https://voronezh.hh.ru/vacancy/70867763?from=v...	ООО ИНТЕРТРАНСАВТОМ
10	HeadHunter	Водитель Яндекс такси	https://voronezh.hh.ru/vacancy/77058566?from=v...	ИП Складчиков Анд МихайловичСанкт-Петерф
11	HeadHunter	Персональный водитель Бизнес такси	https://voronezh.hh.ru/vacancy/76948498?from=v...	ATOMLINE (ИП Ро Влади Анатолевич)Мос
12	HeadHunter	Водитель-курьер на автомобиле компании	https://voronezh.hh.ru/vacancy/69958302?from=v...	LamodaМос
13	HeadHunter	Водитель на автомобиль ГАЗель NEXT	https://voronezh.hh.ru/vacancy/76951077?from=v...	ООО ОниксКа.
14	HeadHunter	Водитель-экспедитор	https://voronezh.hh.ru/vacancy/76999751?from=v...	ООО Agro-BravoТаш
15	HeadHunter	Водитель-курьер на автомобиле компании	https://voronezh.hh.ru/vacancy/72155816?from=v...	LamodaМос

	site	title	link	organiza
16	HeadHunter	Водитель- курьер на личном автомобиле	https://voronezh.hh.ru/vacancy/55292968? from=v...	ООО Легкий ужинМи
17	HeadHunter	Курьер- водитель с личным автомобилем	https://voronezh.hh.ru/vacancy/77023756? from=v...	5P.LabКиров (Киров обла
18	HeadHunter	Экспедитор	https://voronezh.hh.ru/vacancy/76998122? from=v...	ООО Парфюм ТрейдМи
19	HeadHunter	Курьер- водитель	https://voronezh.hh.ru/vacancy/73798015? from=v...	5P.LabНижний Новго

In []: